Virtualization of Multi-Cell 802.11 Networks: Association and Airtime Control

Mahsa Derakhshani[†], Xiaowei Wang[*], Tho Le-Ngoc[*], Alberto Leon-Garcia[†]
[†]Department of Electrical & Computer Engineering, University of Toronto, Toronto, ON, Canada
[*]Department of Electrical & Computer Engineering, McGill University, Montreal, QC, Canada
Email: mahsa.derakhshani@utoronto.ca; xiaowei.wang@mail.mcgill.ca;
tho.le-ngoc@mcgill.ca; alberto.leongarcia@utoronto.ca

Abstract—This paper investigates the virtualization and optimization of a multi-cell WLAN. We consider the station (STA)-access point (AP) association and airtime control for virtualized 802.11 networks to provide service customization and fairness across multiple internet service providers (ISPs) sharing the common physical infrastructure and network capacity. More specifically, an optimization problem is formulated on the STAs' transmission probabilities to maximize the overall network throughput, while providing airtime usage guarantees for the ISPs. Subsequently, an algorithm to reach the optimal solution is developed by applying monomial approximation and geometric programming iteratively. Based on the proposed three-dimensional Markov-chain model of the enhanced distributed channel access (EDCA) protocol, the detailed implementation of the optimal transmission probability is also discussed. The accuracy of the proposed Markov-chain model and the performance of the developed association and airtime control scheme are evaluated through numerical results.

Index Terms—Wireless local area networks (WLANs), IEEE 802.11e standard, STA-AP association, airtime control, throughput, fairness, Markov chain, geometric programming.

I. Introduction

A. Background and Motivation

Wireless virtualization has recently emerged as an architectural choice for the wireless networks, in which different service providers can share physical infrastructure and wireless resources. Allowing resources to be shared, vitualization can facilitate a flexible and lower-complexity solution to support customized services with finer control over quality-of-service (QoS) features. To enable service customization, the key issue would be to provide isolation among virtual networks running by different service providers. Such isolation can be achieved through contention-free resource allocation techniques based on TDMA and/or FDMA, by isolating resources across service providers. But, in systems using contention-based access protocols, providing isolation among virtual networks is more challenging.

For instance, in virtualized 802.11 WLANs, transmissions of different virtual WLANs (V-WLANs) are closely coupled, although administrative virtualization (i.e., one physical AP advertises multiple service set identifiers (SSIDs)) can already differentiate groups of flows. With a carrier sense multiple access (CSMA)-based MAC, unavoidable collisions act to couple the transmissions of different V-WLANs. Moreover, since the network capacity is shared yet constrained, the increase of traffic in one V-WLAN may reduce the available network capacity to another [1]. Thus, an efficient resource allocation among V-WLANs is essential to manage the MAC-layer couplings.

In a 802.11 WLAN with densely deployed APs, before a STA can access the network, it needs to make a decision about which AP to associate with. In virtualized 802.11 networks, such STA-AP association control could create an opportunity to provide fairness guarantees among different ISPs. In addition to STA-AP association, the airtime among STAs associated to the same AP can optimally be adjusted. Controlling airtime usage of the STAs provides the opportunity to optimize the WLAN performance (e.g., improving WLAN throughput by exploiting multi-user diversity) as well as another degree of freedom to guarantee fairness among the ISPs.

The main focus of this work is on the virtualization and optimization of a multi-cell WLAN. More specifically, STA-AP association and airtime control are jointly explored to provide fairness and throughput guarantees for different V-WLANs. Taking into account STA transmission rates and ISP airtime reservations, an optimization problem is formulated to adjust the transmission probability of each STA at each AP. The objective is to maximize the overall network throughput, while keeping a total airtime guarantee for each ISP. To solve the formulated problem which is non-convex and thus computationally intractable, an iterative algorithm is developed through successive geometric programming. This algorithm can achieve an optimal solution with an affordable complexity.

Furthermore, to implement the optimal transmission probabilities, we model the EDCA protocol with a three-dimensional Markov chain and establish the relationship between the transmission probability of a STA and the EDCA parameters. Based on this relationship, we developed a control algorithm to approach the optimal transmission probability by jointly manipulating the EDCA parameters such as contention window size (CW) and arbitration inter-frame space (AIFS). Finally, through numerical results, we verify the performance enhancement provided by the developed STA-AP association and airtime control approach in terms of throughput and fairness.

B. Related Works

1) STA-AP Association and Airtime Control: In most current vendor implementations, 802.11 STAs choose the AP with the highest received signal-to-noise ratio (SNR) to connect with. Since the STA density is often uneven in the network [2], [3], the Max-SNR approach can lead to an unbalanced distribution of STAs among APs, causing unfairness. In order to balance the load of APs, several STA-AP association algorithms have been presented in the literature, mostly by maximizing the minimum throughput of all STAs [4]–[7]. Nevertheless, in a basic service set (BSS) including an AP and its associated multi-rate STAs, it is shown that the throughput is limited by the STA with the lowest data rate. This phenomenon is also to as the *performance anomaly* problem [8]. Thus, comparing with the Max-SNR approach, these load-balancing approaches improve the max-min fairness among STAs at the cost of decreasing the aggregate throughput.

To address the *performance anomaly* and balance the trade-off between aggregate throughput and fairness, proportional fair throughput allocation has widely been considered in multi-rate 802.11 WLANs [9]–[13]. In [9], proportional fairness is studied in a single BSS. It is shown that *propotional fairness* leads to an *airtime-fairness*, where equal airtime usage is provided to all STAs. Moreover, in a multi-AP WLAN, [11], [12] study STA-AP association problem with an objective to maximize the proportional fairness. More precisely, association control is implemented in a form of airtime allocation, where the transmission time of STAs at different APs are jointly optimized [11], [12].

Precise controlling of the STA airtime usage in a 802.11 contention-based WLAN is very difficult due to the distributed and random nature of their CSMA-based MAC protocol. Nevertheless, the airtime control algorithms have been proposed in the literature by manipulating the MAC parameters [14]–[17]. For instance, [15] discusses how to dynamically modify the CW based on the transmission data rate and estimated packet error rate (PER) in a WLAN aiming to guarantee airtime fairness. Similarly, in [16], CW is adjusted according to the traffic load of each STA to improve throughput and fairness in the WLAN. Furthermore, [17] proposes a machine-learning approach to adjust AIFS and CW aiming to achieve airtime fairness and reduce frame delay. Such heuristic works provide pretty good understanding of the influence of specific MAC parameters such as CW and AIFS on the transmission probability and thus airtime usage of the STAs. However, the discussions are limited to adjusting the parameters separately and thus can only find a suboptimal solution in a subset of the feasibility region.

In a virtualized WLAN serving multiple ISPs, STA-AP association and airtime control become more challenging. The reason is that fairness guarantees and service customization are required for ISPs, while there are unavoidable couplings among the STA transmissions of different ISPs in the network. There are a few works addressing only airtime control in the literature. Considering a virtualized single-cell WLAN, in [18],

a heuristic airtime control algorithm is proposed to achieve the target airtime usage for each ISP by controlling the minimum CW of each STA. Similarly, [19] addresses optimizing CW using control theory. But, as the discussion is limited to controlling minimum CW, the optimality of the result might be sacrificed. In a multi-cell WLAN, [1] presents an analysis on the feasibility region of ISP airtimes and characterizes the ISP airtimes at the rate region boundary. Furthermore, a distributed algorithm is developed to allocate airtime slices among ISPs and flow rates within each slice in a max-min fair manner. Since max-min fairness is used as an objective for rate allocation among the flows in each ISP, the optimality of the achieved total throughput may not be guaranteed in a multi-rate WLAN. In addition, the association control is not discussed in [1].

2) IEEE EDCA Modeling and Optimization : To implement the optimal STA-AP association and airtime allocation, it is desirable to control the transmission probability of each STA. However, in a CSMA/CA-based WLAN, the only controllable parameters are the MAC layer parameters. Thus, it is essential to mathematically model the effects of such parameters on the transmission probabilities of STAs.

The EDCA modeling has received considerable attention in the literature. There are several studies, such as [20], [21], that only focus on numerically solving the stationary state (transmission probability of each STA) given the network configuration. Another group of works, e.g., [22]–[29], provide models for the EDCA protocol with the help of a Markov chain, inspired by the classical work of Bianchi [30]. Using such Markov chain models, it is possible to optimize the network by tuning the EDCA parameters. However, due to the complexity of the proposed EDCA models, it is hard to establish an explicit relationship between the transmission probabilities and the controllable EDCA parameters. Thus, these works mostly provide a numerical (e.g., [22]–[25]) or approximate analytical solution in terms of MAC parameters (e.g., [26]–[29]). Moreover, none of these works has jointly controlled the EDCA parameters.

Our proposed Markov chain model is based on [31], with which an explicit relationship between the transmission probability and EDCA MAC parameters can be established. One limitation of the EDCA model proposed in [31] is the accuracy of AIFS differentiation. Unlike in DCF, the time duration each STA has to wait before its backoff process (i.e. AIFS) is different in EDCA. This can lead to the result that the number of contending STAs is not time-homogeneous [21]. Therefore, the collision probability that each STA faces is also not time-homogeneous. But, as pointed out in [27], time-homogeneity assumption greatly simplifies the modeling complexity, and thus, the steady state performance can be characterized as explicit functions of backoff parameters. Furthermore, the accuracy of such model can be effectively improved by setting a relatively large initial CW.

C. Structure

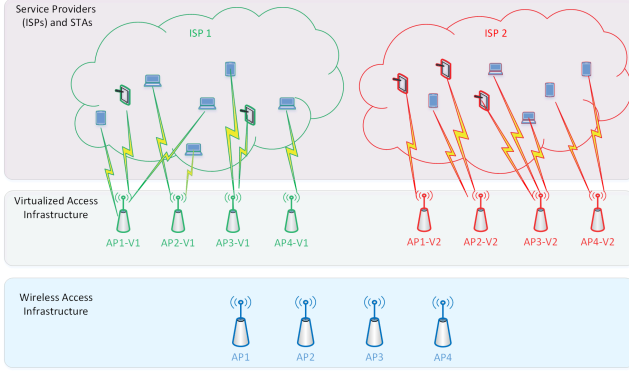The rest of this paper is organized as follows. Section II
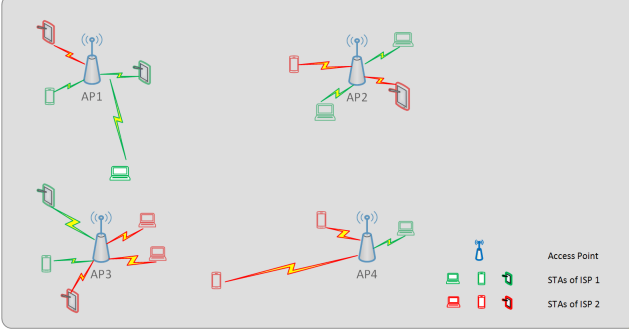
Fig. 1: Layered system model

Fig. 2: System model

presents an overview of the system configuration and modeling. In Section III, we first analyze the feasibility region of the transmission probabilities based on the proposed Markov chain model for IEEE EDCA. Then, we formulate the transmission probability optimization problem, which maximizes the system throughput and guarantees the fairness among the ISPs. In Section IV, the implementation details of MAC parameter control are discussed in order to achieve the optimal transmission probability. Illustrative results are provided in Section V to evaluate the performance of the developed algorithms. Section VI provides some concluding remarks.

II. System Configuration and Modeling

We consider an IEEE 802.11-based WLAN that consists of a large number of APs. APs operate on non-overlapping frequency channels. Let $\mathcal{A}$ be the set of APs and $N_a = |\mathcal{A}|$ be the total number of APs. Each AP has a limited coverage area and all STAs are randomly distributed in the field. The network carries traffic belonging to a number of different ISPs (also referred to as V-WLANs). Let $\mathcal{K}$ be the set of ISPs using the network. Furthermore, let $\mathcal{S}_k$ be the set of STAs of ISP $k \in \mathcal{K}$ and $N_k = |\mathcal{S}_k|$ be the number of STAs belong to ISP k. Furthermore, let $\mathcal{S}$ be the set of all STAs and $N_s = \sum_{k \in \mathcal{K}} N_k$ be the total number of STAs in the network. The network is administratively virtualized, i.e., each AP will broadcast multiple different SSIDs, one for each ISP. Figure 1 illustrates an example of the network architecture with four physical APs and two ISPs.
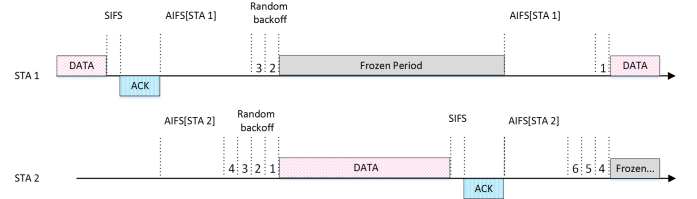

Fig. 3: IEEE 802.11e EDCA channel access procedure

A. Enhanced Distributed Channel Access (EDCA)

To access the channel, the STAs are assumed to follow the EDCA MAC protocol of 802.11e standard. EDCA protocol is designed to enhance the basic MAC mechanism, i.e., distributed coordination function (DCF), aiming to support service differentiation. Similar to DCF, EDCA is also a contention-based access scheme, based on CSMA/CA using binary exponential backoff rules to manage retransmission of collided packets. Here, we briefly review the EDCA operation, as standardized by 802.11 protocol [30], [32], [33].

EDCA requires a STA to monitor the channel before any transmission. If the channel is sensed idle for a time interval equal to an arbitration inter-frame space (AIFS), the STA transmits. Otherwise, if the STA senses a transmission either immediately or during the AIFS, it continues monitoring the channel. When the channel is measured idle for an AIFS, the STA backoffs for a random period of time. EDCA uses a discrete-time backoff mechanism, i.e., the time following an AIFS is slotted. The backoff time is selected according to a uniform distribution in the interval $[0, W_i]$ where W_i represents the contention window of STA $i \in \mathcal{S}$ [30], [32], [33].

STA i starts a packet transmission at backoff stage 0, where W_i is set equal to the minimum contention window size $W_{\min,i}$. Then, after each unsuccessful transmission, STA i moves to the next backoff stage and doubles W_i. When the maximum backoff stage m_i is reached, W_i is no longer increased and stays at $2^{m_i} W_{\min,i}$. If the STA experiences a collision at the backoff stage m_i, it will retry transmission for at most h_i times, where h_i is the retransmission limit at the maximum backoff stage. If the packet is still not successfully transmitted after h_i retransmissions, it will be discarded [30], [32], [33].

The backoff time counter is decremented and a STA transmits when the backoff counter reaches zero. If the data frame is successfully received, the AP waits for a period of time called short inter-frame space (SIFS) and then sends an acknowledgment (ACK). Once the channel is sensed busy while counting down, the STA will freeze its backoff counter and continue decrementing when channel finds idle again. Figure 3 illustrates an example of the channel-access procedure of two STAs using EDCA [30], [32], [33].

EDCA allows to have variable waiting times and contention parameters (i.e., AIFS and $W_{\min}$) to enable service differentiation. Using EDCA, for each transmission opportunity, a STA can transmit multiple back-to-back packets for a fixed period

of time. Let T_{TXOP} be the duration of a data frame. Then, the duration of a successful transmission becomes

$$T_s = T_{\text{TXOP}} + \text{SIFS} + \gamma + \text{ACK} + \gamma + \text{AIFS} \tag{1}$$

where γ denotes the transmission delay. Similarly, the duration of a collision can be calculated as

$$T_c = T_{\text{TXOP}} + \gamma + \text{AIFS}. \tag{2}$$

Note that when the colliding STAs use different AIFS values, the AIFS value in (2) should take the value of the largest AIFS. But, since the fixed transmission duration T_{TXOP} (in the order of ms) is much larger than AIFS in the order of μs, the difference between the AIFS can be ignored here.

B. Association Control via Transmission Probabilities

In a WLAN with APs densely deployed, STAs need to determine which APs to connect with. We aim to generalize the association control problem by adjusting the transmission probability of each STA at any AP, rather than selecting one AP to associate with. Thus, we define τ_i^a ($0 \leq \tau_i^a \leq 1$) as the probability that STA i attempts to transmit at AP a in a general time-slot. Consequently, the probability that a time-slot is idle in the BSS including AP a is

$$P_{\text{idle}}^a = \prod_{i \in \mathcal{S}}(1 - \tau_i^a). \tag{3}$$

In a given BSS, transmitted packets will be received successfully, if exactly one STA transmits on the channel. Thus, the probability of a successful transmission initiated by STA i becomes

$$P_{\text{succ},i}^a = \tau_i^a \prod_{i' \in \mathcal{S}, i' \neq i}(1 - \tau_{i'}^a). \tag{4}$$

Since ACK and SIFS are relatively small (in the order of μs) compared with T_{TXOP} (in the order of ms), we approximate T_s and T_c to be of the same size and denote them by T. Consequently, the expected length of a general time-slot becomes

$$\mathbb{E}\{T_g\} = \delta P_{\text{idle}}^a + (1 - P_{\text{idle}}^a)T \tag{5}$$

where δ is the duration of an idle time-slot. Furthermore, the expected information (in bits) transmitted by STA i to AP a in a general time-slot can be derived as

$$\mathbb{E}\{I_g\} = P_{\text{succ},i}^a r_i^a T_{\text{TXOP}} \tag{6}$$

where r_i^a represents the transmission data rate of the link between STA i and AP a. As defined in [30], based on the (5) and (6), the throughput of STA i at AP a becomes

$$T_i^a = \frac{\mathbb{E}\{I_g\}}{\mathbb{E}\{T_g\}} = \frac{P_{\text{succ},i}^a r_i^a T_{\text{TXOP}}}{P_{\text{idle}}^a \delta + (1 - P_{\text{idle}}^a)T}. \tag{7}$$

Let define a new variable $x_i^a = \frac{\tau_i^a}{1 - \tau_i^a}$ ($x_i^a \geq 0$), which represents the expected number of consecutive transmission attempts by STA i at AP a as [1], [9]. Consequently, P_{idle}^a and $P_{\text{succ},i}^a$ will be transformed into

$$P_{\text{idle}}^a = \frac{1}{\prod_{i \in \mathcal{S}}(1 + x_i^a)}, \tag{8}$$

$$P_{\text{succ},i}^a = \frac{x_i^a}{\prod_{i' \in \mathcal{S}}(1 + x_{i'}^a)} = x_i^a P_{\text{idle}}^a. \tag{9}$$

Subsequently, from (8) and (9), T_i^a can be represented in terms of x_i^a as

$$T_i^a = \frac{x_i^a P_{\text{idle}}^a r_i^a T_{\text{TXOP}}}{T - (T - \delta)P_{\text{idle}}^a} = \frac{x_i^a r_i^a t}{\prod_{i' \in \mathcal{S}}(1 + x_{i'}^a) - t'} \tag{10}$$

where $t = \frac{T_{\text{TXOP}}}{T}$ and $t' = \frac{T - \delta}{T}$.

In addition to the throughput of each STA, the fraction of time that each STA spends for transmission could be considered as another performance metric, specifically in order to measure and preserve fairness among different STAs or ISPs. The total access airtime for STA i–including both successful transmissions and collisions–becomes

$$T_{\text{air},i}^a = \frac{P_{\text{coll},i}^a T + P_{\text{succ},i}^a T}{P_{\text{idle}}^a \delta + (1 - P_{\text{idle}}^a)T} \tag{11}$$

where $P_{\text{coll},i}^a = \tau_i^a \left[1 - \prod_{i' \in \mathcal{S}, i' \neq i}(1 - \tau_{i'}^a)\right]$ is the probability that STA i suffers from a collision in a general time-slot as defined in [9]. Consequently,

$$T_{\text{air},i}^a = \frac{\tau_i^a}{1 - P_{\text{idle}}^a t'} = \frac{x_i^a \prod_{i' \in \mathcal{S}, i' \neq i}(1 + x_{i'}^a)}{\prod_{i' \in \mathcal{S}}(1 + x_{i'}^a) - t'}. \tag{12}$$

In this work, we aim to maximize the overall network throughput, while guaranteeing a minimum requirement on the aggregate airtime of each ISP. To this end, the transmission probability of STAs (τ_i^a) need to be adaptively optimized by maximizing the aggregate throughput of all STAs at all APs (i.e.,$\sum_{i \in \mathcal{S}, a \in \mathcal{A}} T_i^a$). Furthermore, for each ISP (e.g., ISP k), a constraint needs to be set in order to keep the total airtime of all STAs belonging to ISP k larger than a minimum requirement. More specifically, $\sum_{i \in \mathcal{S}_k, a \in \mathcal{A}} T_{\text{air},i}^a \geq \eta_k$ where η_k denotes the target share of the airtime for ISP k.

Accordingly, to formulate such optimization problem, the feasibility region of τ_i^a (or x_i^a) is required. Thus, we study the behavior of a single STA -which is using EDCA- with a three-dimensional Markov chain. With the aid of the proposed Markov model, we can learn how to implement or control τ_i^a in terms of EDCA parameters. As a result, based on the established relationship between τ_i^a and EDCA parameters, we would be able to analyze its feasibility region and also design an algorithm to control EDCA parameters to approach the optimal τ_i^a.

III. OPTIMIZATION PROBLEM

In this section, we present the STA-AP association and airtime control optimization problem based on the system model introduced in Section II. We first use a Markov chain model to model the EDCA protocol. With this model, we can analyze the feasibility region of τ_i^a for each STA. Then, the optimization problem is formed and solved by applying monomial approximation and geometric programming.

$$\tau_i = \sum_{j=0}^{m_i+h_i} b_{j,0,0} = \frac{1-p_i^{m_i+h_i+1}}{1-p_i}\left[L_i\frac{1-q_i}{q_i} + \frac{1+p_iN}{p_i}\frac{1-(1-p_i)^{A_i+1}}{(1-p_i)^{A_i+1}} + \frac{1-p_i^{m_i+h_i+1}}{1-p_i} + \right.$$

$$\left. \frac{1+Np_i}{2(1-p_i)^{A_i}}\left(W_{\min,i}\left[\frac{1-(2p_i)^{m_i+1}}{1-2p_i} + \frac{2^{m_i}p_i^{m_i+1}(1-p_i^{h_i})}{1-p_i}\right] - \frac{1-p_i^{m_i+h_i+1}}{1-p_i}\right)\right]^{-1} \qquad (13)$$
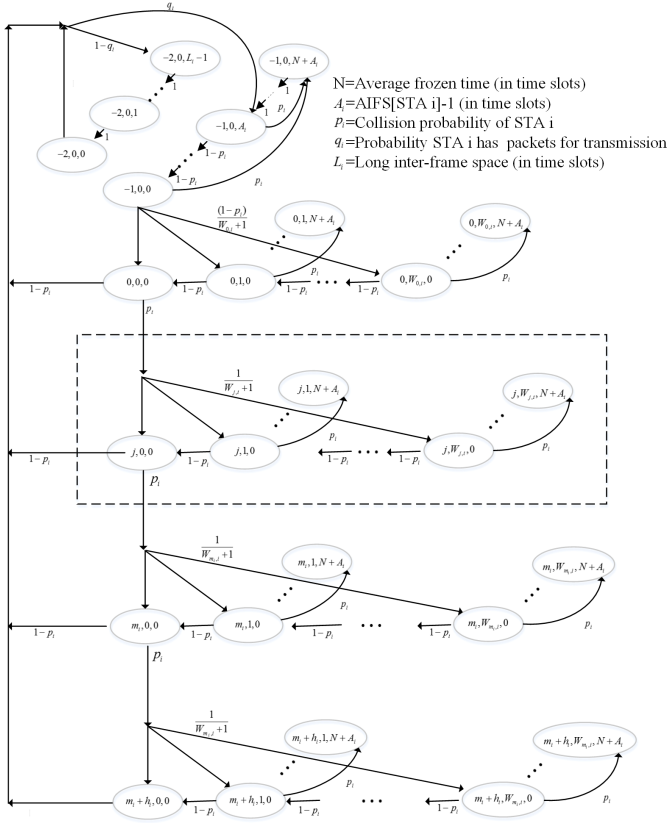

N=Average frozen time (in time slots)
A_i=AIFS[STA i]-1 (in time slots)
p_i=Collision probability of STA i
q_i=Probability STA i has packets for transmission
L_i=Long inter-frame space (in time slots)

Fig. 4: Three dimensional Markov chain model for one STA

A. Markov Chain Model for EDCA

To study the feasibility region of τ_i^a and its implementation in the EDCA MAC protocol, we establish a three-dimensional Markov chain. Such model enables us to estimate the transmission probability of the STAs in a WLAN employing the EDCA protocol. Different from the previous works, which discuss a single parameter in the EDCA protocol (e.g., [15]), we aim at developing a general mathematical model that can show the influences of all the EDCA parameters on the transmission probability.

Figure 4 shows the three-dimensional Markov chain model to describe the behavior of a single STA, which uses the EDCA MAC protocol. Since STA transmissions are only coupled within a BSS, we consider a single BSS where an AP and its associated STAs reside in. Thus, in this subsection without loss of generality, we remove the index of a (e.g., τ_i^a is replaced by τ_i) to keep the notation simple.

Our proposed three-dimensional Markov chain model is an

extension to the presented model in [31]. To be able to further control the competition among STAs and thus the collision probability in a BSS, we introduce two new MAC parameters (i.e., q_i and L_i for STA i) for each STA. Similar to [34], after a successful transmission or a packet drop, the STA has to flip a biased coin with successful probability q_i to enter the backoff process. Whenever the STA fails to enter the backoff process, it needs to wait for a period of long inter-frame space (i.e., L_i time-slots) before another try. These two variables are incorporated in the proposed Markov chain.

It should be noted that time in this Markov chain is slotted and the interval between any two adjacent states is a general time-slot. Here, a general time-slot can be an idle time-slot δ, a successful transmission duration (T_s) or a collision duration (T_c). In this Markov chain, the triple $\{s(t), b(t), v(t)\}$ denotes the state of the STA at time t. More specifically, $s(t)$ represents the backoff/retransmission stage of STA, where $-2 \leq s(t) \leq m_i + h_i$ ($s(t) = -2, -1$ will be explained later). Furthermore, $b(t)$ denotes the backoff counter, which takes a value between 0 and the contention window of the current backoff stage. Furthermore, $v(t)$ represents the remaining time of a frozen period or the waiting time to the next trial when the STA already failed to enter the backoff process.

The detailed transition probabilities between states of the Markov chain and the stationary condition of the Markov chain can be found in the Appendix. With such stationary conditions, the transmission probability τ of each STA is derived as function of the detailed EDCA parameters as in (13), where p_i is the collision probability faced by STA i as defined in (26).

B. Feasibility Region Analysis

In this subsection, we study the feasibility region of τ_i^a in order to complete the optimization problem formulation. Based on (13), the adjustable variables to tune τ_i^a are m_i^a, h_i^a, L_i^a, q_i^a, A_i^a, and $W_{j,i}^a$ where $0 \leq j \leq m_i^a + h_i^a$. Recall that $0 \leq q_i^a \leq 1$, $0 \leq L_i^a$, and $0 \leq W_{j,i}^a$. Furthermore, since $A_i = \text{AIFS}[\text{STA } i] - 1$, and AIFS $>$ SIFS in EDCA, we have $A_i \geq$ SIFS. Assuming that SIFS is equal to one time-slot, we have $A_i^a \geq 1$ without loss of generality.

1) Lower Bound of τ_i^a: Let $\underline{\tau}_i^a$ be the lower bound of τ_i^a. From (13), it is clear that $\tau_i^a \to 0$ when $L_i^a \to \infty$, or $q_i^a \to 0$, or $A_i^a \to \infty$. Consequently,

$$\underline{\tau}_i^a = 0 \qquad (14)$$

2) Upper Bound of τ_i^a: Let $\bar{\tau}_i^a$ be the upper bound of τ_i^a. To calculate $\bar{\tau}_i^a$, we need to make the denominator of τ_i^a in

(13) as small as possible. Thus, we set $L_i^a = 0$ (or $q_i^a = 1$) and $W_{\min,i}^a = 0$, which make the first and fourth terms in the denominator zero. Then, by setting A_i to its lower bound (i.e., $A_i^a = 1$), we can minimize the second term in the denominator of τ_i^a. Consequently, at $L_i^a = 0$, $W_{\min,i}^a = 0$, and $A_i^a = 1$,

$$\tau_i^a = \left[1 + \frac{(1 + p_i^a N)(2 - p_i^a)}{(1 - p_i^a)(1 - (p_i^a)^{m_i + h_i + 1})}\right]^{-1} \quad (15)$$

From (15), it is clear that the upper bound of τ_i^a can be achieved when $m_i^a + h_i^a \to \infty$. Thus, $\bar{\tau}_i^a$ can be written as

$$\bar{\tau}_i^a = \left[1 + \frac{(1 + p_i^a N)(2 - p_i^a)}{1 - p_i^a}\right]^{-1} \quad (16)$$

C. Optimization Problem

Taking into account the feasibility region of τ_i^a derived in Section III-B, here, we can mathematically present the transmission probability optimization problem. The objective is to maximize the overall network throughput, while distributing access airtime among different ISPs according to their reservations. More specifically, the optimization can be formulated as

$$\max_{\boldsymbol{X}, \boldsymbol{P}} \sum_{i \in \mathcal{S}, a \in \mathcal{A}} \frac{x_i^a r_i^a t}{\prod_{i' \in \mathcal{S}}(1 + x_{i'}^a) - t'}, \quad \text{subject to,} \quad (17a)$$

$$\sum_{i \in \mathcal{S}_k, a \in \mathcal{A}} \frac{x_i^a \prod_{i' \in \mathcal{S}, i' \neq i}(1 + x_{i'}^a)}{\prod_{i' \in \mathcal{S}}(1 + x_{i'}^a) - t'} \geq \eta_k, \ \forall k \in \mathcal{K} \quad (17b)$$

$$\frac{x_i^a}{1 + x_i^a} \leq \bar{\tau}_i^a, \ \forall i \in \mathcal{S}, a \in \mathcal{A} \quad (17c)$$

$$\bar{\tau}_i^a \left(1 + \frac{(1 + p_i^a N)(2 - p_i^a)}{1 - p_i^a}\right) = 1, \ \forall i \in \mathcal{S}, a \in \mathcal{A} \quad (17d)$$

$$p_i^a = 1 - \prod_{i' \in \mathcal{S}, i' \neq i} 1 - \frac{x_{i'}^a}{1 + x_{i'}^a}, \ \forall i \in \mathcal{S}, a \in \mathcal{A}. \quad (17e)$$

where $\boldsymbol{X} = [x_i^a]$ ($x_i^a \geq 0$) and $\boldsymbol{P} = [p_i^a]$ ($0 \leq p_i^a \leq 1$). Let us recall that $x_i^a = \frac{\tau_i^a}{1 - \tau_i^a}$ and $p_i^a = 1 - \prod_{i' \neq i}(1 - \tau_{i'}^a)$. It should be noted that the optimization problem is alternatively formulated with respect to x_i^a instead of τ_i^a, since it will prove useful to solve the problem.

In the preceding optimization problem, the objective function in (17a) represents the overall network throughput (i.e., $\sum_{i \in \mathcal{S}, a \in \mathcal{A}} T_i^a$) based on (10). Furthermore, constraints in (17b) guarantee the minimum airtime reservations for all ISPs (i.e., $\sum_{i \in \mathcal{S}_k, a \in \mathcal{A}} T_{\text{air},i}^a \geq \eta_k$) based on (12). This set of constraints enable controlling ISPs' share of access airtime regardless of their number of STAs. To guarantee that the provided solution fits into the feasibility region, in constraint (17c), τ_i^a is limited to its upper bound. Then, the equality constraint (17d) establishes the relationship between the upper bound $\bar{\tau}_i^a$ and p_i^a based on (16). Finally, constraint (17e) establishes the relationship between p_i^a and all $x_{i'}^a$, $\forall i' \in \mathcal{S}$, $i' \neq i$ according to (26).

The formulated problem is non-convex and thus intractable to solve. However, it potentially looks like an extension of Geometric Programming (GP) (defined in Section VII-B).

Thus, by applying successive transformation strategies, we will try to convert the original problem into a series of standard GP problems that can be solved to reach an optimal solution. First, we introduce three auxiliary variables, $y^a = \prod_{i' \in \mathcal{S}}(1 + x_{i'}^a) - t'$, $\forall a \in \mathcal{A}$, $u_i^a = 1 - p_i^a$, $\forall i \in \mathcal{S}, \forall a \in \mathcal{A}$, and $t_i^a = 1 + x_i^a$, $\forall i \in \mathcal{S}, \forall a \in \mathcal{A}$. Then, the optimization problem in (17) can be transformed into

$$\min_{\boldsymbol{X}, \boldsymbol{T}, \boldsymbol{U}, \boldsymbol{P}, \boldsymbol{Y}} \sum_{i \in \mathcal{S}, a \in \mathcal{A}} -\frac{x_i^a r_i^a t}{y^a}, \quad \text{subject to,} \quad (18a)$$

$$\frac{\prod_{i \in \mathcal{S}}(1 + x_i^a)}{t' + y^a} = 1, \ \forall a \in \mathcal{A} \quad (18b)$$

$$\frac{\eta_k + 1}{1 + \sum_{i \in \mathcal{S}_k, a \in \mathcal{A}} \frac{x_i^a \prod_{i' \neq i} t_i^a}{y^a}} \leq 1, \ \forall k \in \mathcal{K} \quad (18c)$$

$$\frac{x_i^a}{1 + x_i^a} \leq \frac{1}{1 + \frac{(1 + p_i^a N)(2 - p_i^a)}{1 - p_i^a}}, \ \forall i \in \mathcal{S}, a \in \mathcal{A} \quad (18d)$$

$$u_i^a \prod_{i' \in \mathcal{S}, i' \neq i} t_{i'}^a = 1, \ \forall i \in \mathcal{S}, a \in \mathcal{A} \quad (18e)$$

$$\frac{t_i^a}{1 + x_i^a} = 1, \ \forall i \in \mathcal{S}, a \in \mathcal{A} \quad (18f)$$

$$u_i^a + p_i^a = 1, \ \forall i \in \mathcal{S}, a \in \mathcal{A} \quad (18g)$$

where $\boldsymbol{T} = [t_i^a]$ ($t_i^a \geq 1$), $\boldsymbol{U} = [u_i^a]$ ($0 \leq u_i^a \leq 1$), $\boldsymbol{Y} = [y^a]$ ($y^a > 0$). Nevertheless, the transformed problem is not still in a GP form. One reason is that the objective function (18a) is not a posynomial because of negative multiplicative coefficients. To deal with such problem, first, we equivalently substitute the objective function by $\sum_{i \in \mathcal{S}, a \in \mathcal{A}} -x_i^a r_i^a t(y^a)^{-1} + M$ where M is a sufficiently large positive constant. Adding M makes sure that the objective function is always positive. Then, we introduce an additional auxiliary variable $x_0 \geq 0$. By minimizing x_0 and guaranteeing constraint C11 in (19), we can effectively minimize the objective function in (18a). Consequently,

$$\min_{\boldsymbol{X}, \boldsymbol{T}, \boldsymbol{U}, \boldsymbol{P}, \boldsymbol{Y}, x_0} x_0, \quad \text{subject to,} \quad (19a)$$

$$\text{C11} : \frac{M}{x_0 + \sum_{i \in \mathcal{S}, a \in \mathcal{A}} \left(\frac{x_i^a r_i^a t}{y^a}\right)} \leq 1$$

$$\text{C12} : \frac{\prod_{i \in \mathcal{S}}(1 + x_i^a)}{t' + y^a} = 1, \ \forall a \in \mathcal{A}$$

$$\text{C13} : \frac{\eta_k + 1}{1 + \sum_{i \in \mathcal{S}_k, a \in \mathcal{A}} \frac{x_i^a \prod_{i' \in \mathcal{S}, i' \neq i} t_{i'}^a}{y^a}} \leq 1, \ \forall k \in \mathcal{K}$$

$$\text{C14} : \frac{u_i^a x_i^a + (1 + N) x_i^a}{u_i^a + N(u_i^a)^2 x_i^a} \leq 1, \ \forall i \in \mathcal{S}, a \in \mathcal{A}$$

$$\text{C15} : u_i^a \prod_{i' \in \mathcal{S}, i' \neq i} t_{i'}^a = 1, \ \forall i \in \mathcal{S}, a \in \mathcal{A}$$

$$\text{C16} : \frac{t_i^a}{1 + x_i^a} = 1, \ \forall i \in \mathcal{S}, a \in \mathcal{A}$$

In the preceding optimization problem, p_i^a is replaced by $1 - u_i^a$ based on constraint in (18g). The optimization problem in (19) belongs to the class of *complementary GP* problems

Algorithm 1 : GP-based Association Control Algorithm

Initialize x_i^a, t_i^a, p_i^a, u_i^a, for all $i \in \mathcal{S}, a \in \mathcal{A}$, y_a for all $a \in \mathcal{A}$, x_0;
Record the current system state as $\boldsymbol{Z} = (\boldsymbol{X}, \boldsymbol{T}, \boldsymbol{U}, \boldsymbol{P}, \boldsymbol{Y}, x_0)$;

repeat
    Compute the ratio α of each monomial term in the denominator of C11, C12, C13, C14, and C16 according to (30), at the current system state $\boldsymbol{Z}$;
    Apply monomial approximation to the denominators mentioned above according to (29);
    Solve the resulting GP problem using cvx;
    Update the current system state $\boldsymbol{Z} = (\boldsymbol{X}, \boldsymbol{T}, \boldsymbol{U}, \boldsymbol{P}, \boldsymbol{Y}, x_0)$;
until all x_i^a converge.

Compute the optimal transmission probabilities $\tau^{*a}_i = \frac{x^{*a}_i}{1 + x^{*a}_i}$

---

that allow upper bound constraints on the ratio between two posynomials and equality constraints on the ratio between a monomial and a posynomial [35], [36]. By approximating the posynomials in the denominator of such constraints, a *complementary GP* can be turned into a standard form of GP. Consequently, the optimal solution can be achieved by iteratively applying monomial approximations and solving a series of GPs. The arithmetic-geometric mean inequality can be used to approximate a posynomial with a monomial. The details of such monomial approximation are provided in Section VII-B.

Accordingly, we propose an iterative algorithm to reach to an optimal solution of the transmission probability optimization problem. In each iteration, monomial approximations are applied to the denominator of C11, C12, C13, C14, and C16. Then, the resulting GP can be solved for instance by using a standard interior-point algorithm. More specifically, Algorithm 1 presents different steps need to be performed until convergence.

D. Asymptotic Complexity Analysis

In this subsection, we study the asymptotic complexity and scalability of Algorithm 1 in terms of the number of STAs (i.e., N_s) and the number of APs (i.e., N_a).

In Algorithm 1, for each iteration, the computational complexity is incurred by applying monomial approximations and solving the resulting GP problem. Suppose C_{MA} and C_{GP} denote the required computing efforts for monomial approximations and solving GP problem in each iteration.

More specifically, first in each iteration, the denominators of constraints C11, C12, C13, C14, and C16 need to be approximated according to (29). The required computational complexity for these monomial approximations is proportional to the number of monomial terms in the denominators of constraints C11, C12, C13, C14 and C16. Consequently, the complexity of monomial approximations totals to

$$C_{\mathrm{MA}} = \mathcal{O}(N_a N_s) + \mathcal{O}(N_a N_s) + \mathcal{O}(N_a N_s^2)$$
$$+ \mathcal{O}(N_a N_s) + \mathcal{O}(N_a N_s) = \mathcal{O}(N_a N_s^2) \quad (20)$$

Subsequently, the resulting GP problem needs to be solved

by transforming it to a convex problem. It is reported in [37] that the worst-case computational complexity of this approach is $\mathcal{O}(pn^3)$, where n is the number of variables, and p is the total number of terms in all monomials and posynomials in the objective function and constraints. The number of variables of the optimization problem in (19) is

$$n = 4N_a N_s + 2N_a + 1 = \mathcal{O}(N_a N_s) \quad (21)$$

Furthermore, the total number of terms in all monomials and posynomials can be counted as

$$p = 1 + (4N_a N_s + 1) + N_a(N_s + 2)$$
$$+ \sum_k (1 + N_k N_a(N_s + 1))$$
$$+ 7N_a N_s + N_a N_s^2 + 2N_a N_s$$
$$= \mathcal{O}(N_a N_s^2) \quad (22)$$

where each term in (22) is respectively the number of terms in all the monomials and posynomials in the objective function and constraints C11-C16. Based on (21) and (22), to solve the GP problem, the complexity is

$$C_{\mathrm{GP}} = \mathcal{O}(pn^3) = \mathcal{O}(N_a^4 N_s^5) \quad (23)$$

Thus, based on (20) and (23), *per-iteration asymptotic complexity* of Algorithm 1 becomes

$$C_{\mathrm{Alg1-iteration}} = C_{\mathrm{MA}} + C_{\mathrm{GP}} = \mathcal{O}(N_a^4 N_s^5) \quad (24)$$

We now turn to studying the number of iterations (denoted by Δ) required for Algorithm 1 to converge. Figures 5a and 5b illustrate numerical results on Δ versus number of APs (i.e., N_a) and average number of STAs per AP (i.e., λ_{mean}), respectively. The simulation setup used in these figures is the same as the one presented in Section V. In Figure 5a, λ_{mean} is set to $\frac{10}{N_a}$, so that the expected total number of STAs in the WLAN stays fixed. Figure 5a shows that Δ increases with N_a when N_a is small, and then, it becomes *steady* for a larger range of N_a. In Figure 5b, N_a is fixed equal to 4 and λ_{mean} is varying. Figure 5b shows that Δ grows *linearly* over a typical range of λ_{mean}. It should be noted that the fluctuations in Δ are caused by the randomness in the number of STAs per AP, which follows a Poisson distribution with mean of λ_{mean}.

Therefore, based on (24) and the numerical results in Figures 5a and 5b, it is shown that the overall complexity of the proposed algorithm only grows *polynomially* with the number of STAs and APs, compared to an exponential complexity required by direct search methods.

IV. IMPLEMENTATION DETAILS

The optimal transmission probability of STA i at AP a (i.e., τ^{*a}_i) can be obtained from Algorithm 1. But, in the EDCA protocol, the transmission probabilities of STAs are not directly controllable. Instead, what we can control are the MAC-layer parameters (e.g., W_{min} and AIFS) to achieve the optimal performance. In this section, we first verify the accuracy of the relationship between τ_i^a and the EDCA parameters provided in (13) and also the validity of the Markov chain model. Then,
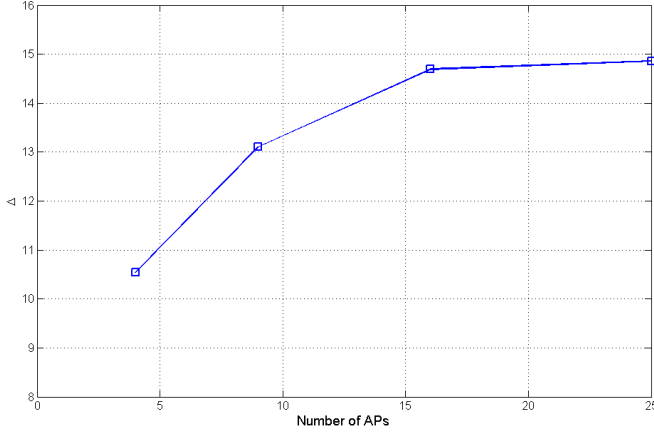
(a) Number of iterations vs. number of APs (N_a)
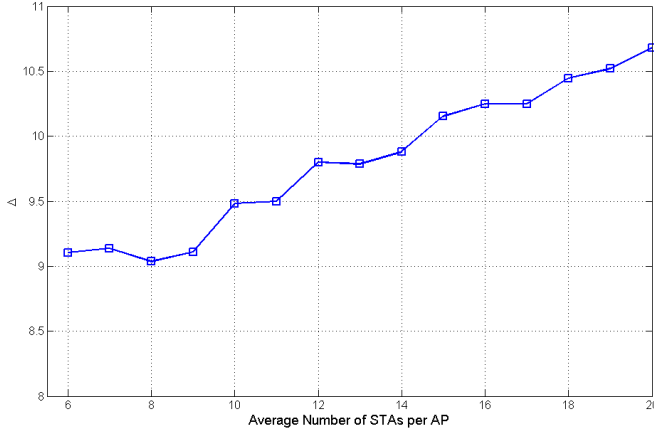
(b) Number of iterations vs. STA density (λ_{mean})

Fig. 5: Number of iterations required for Algorithm 1 to converge

Fig. 6: Accuracy of controlling different parameters

Algorithm 2 : MAC Parameter Control Algorithm

For each STA-AP pair (i, a):
$p_i^a = \prod_{i' \in \mathcal{S}, i' \neq i}(1 - \tau^{*a}_{i'})$
Initialization:
Set $W_{\min,i}^a = 15$, $A_i^a = 6$, $q_i^a = 0.5$, $L_i^a = 100$, $m_i^a = 6$, $h_i^a = 6$;
Compute $W_{\min,i}^a$ from (13), while other variables are fixed;
Round $W_{\min,i}^a$ to an integer;
if $W_{\min,i}^a < 0$ **then**
  Set $W_{\min,i}^a = 0$; Compute L_i^a from (13), while other variables are fixed;
  if $L_i^a < 0$ **then**
    Set $L_i^a = 0$; Compute A_i^a from (13), while other variables are fixed; Round A_i^a to an integer;
    if $A_i^a < 1$ **then**
      Set $A_i^a = 1$; Compute m_i^a from (13), while other variables are fixed; Round m_i^a to an integer;
      if $m_i^a < 0$ **then**
        Set $m_i^a = 0$; Compute h_i^a from (13), while other variables are fixed; Round h_i^a to an integer;
        if $h_i^a < 0$ **then**
          Set h_i^a=0;
Start transmission using EDCA protocol with parameters $\{W_{\min,i}^a, A_i^a, q_i^a, L_i^a, m_i^a, h_i^a\}$.

---

we develop an algorithm to adjust EDCA parameters aiming to implement τ^{*a}_i in the EDCA protocol.

We first investigate the achievable accuracy by controlling different parameters through an example. In this example, we consider one BSS with 6 STAs. The transmission probabilities of STA 2 to STA 6 are fixed (e.g., $\tau_2 = \tau_3 = \tau_4 = \tau_5 = \tau_6 = 0.005$), while the transmission probability of STA 1 (i.e., τ_1) is varied from 0.005 to 0.1.

Figure 6 compares the throughput of STA 1 analytically derived from (10) and numerically measured using an EDCA simulator. In the numerical results, different EDCA parameters are separately adjusted to tune τ_1 to the desired value. From analytical result in Figure 6, it is clear that the throughput of STA 1 increases with its transmission probability τ_1. Regarding the numerical results, the following conclusions can be drawn.

From Figure 6, $W_{\min,i}$ is the best single parameter to be controlled aiming to adjust τ_i to a desired value. A_i and L_i provide a good match and thus can be good choices to tune τ_i. Furthermore, m_i and h_i control is not adequately effective, especially for h_i. Also, as we can observe, the achievable ranges of τ by controlling each parameter (i.e., A_i, $W_{\min,i}$ and L_i) are limited. More specifically, when A_i is being adjusted,
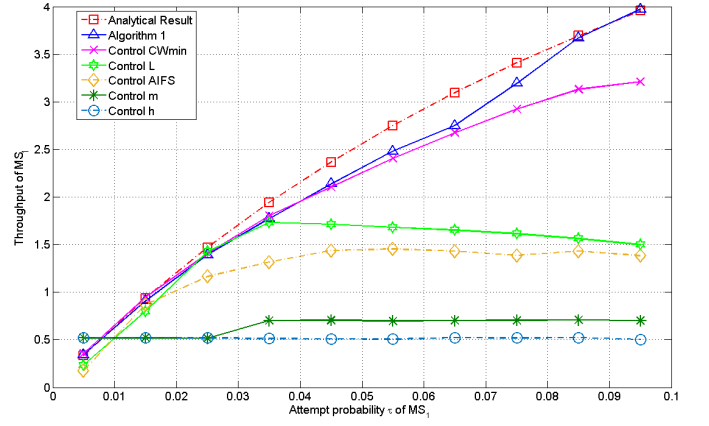
we can learn from Figure 6 that the simulation result matches with the analytical result when τ_1 is in the range $[0.005, 0.15)$. When CWmin is being controlled, the situation is similar, but for a broader range of τ_i ($[0.005, 0.065)$). Figure 6 also suggests that controlling L is effective when τ_1 is in the range $[0.005, 0.035)$. Thus, in summary, to achieve a boarder feasible range of τ, we need to jointly adjust these parameters.

Here, we aim to develop a MAC parameter control algorithm to achieve the optimal transmission probability obtained from the optimization problem in Section III. In Section III-B, the feasibility region of transmission probability is calculated assuming that all the 5 MAC parameters (i.e., $W_{\min}$, A, q or L, m, h) can be freely varied in their feasible ranges. Thus, a control algorithm is required which can be able to adjust all the 5 parameters simultaneously. Thus, an algorithm is proposed taking into account the achievable accuracy by controlling different parameters. Algorithm 2 shows the complete control

Modulation	FEC Rate	Data Rate (Mbps)	SNR (dB)
BPSK	1/2	6	[5,8)
BPSK	3/4	9	[8,10)
QPSK	1/2	12	[10,13)
QPSK	3/4	18	[13,16)
16QAM	1/2	24	[16,19)
16QAM	3/4	36	[19,22)
64QAM	2/3	48	[22,25)
64QAM	3/4	54	[25,∞)

TABLE I: IEEE 802.11a adaptive modulation and coding scheme and the SNR ranges used in the numerical results

procedure.

The achievable throughput for STA 1 in Algorithm 2 is also plotted in Figure 6. As expected, we can learn that Algorithm 2 can approach the desired throughput closely for the total range of τ_1 (i.e., the upper bound τ_1 computed from (16)). Thus, Algorithm 2 can improve the control accuracy compared with controlling the parameters separately.

V. NUMERICAL RESULTS

In this section, we present numerical results to evaluate the performance of the proposed STA-AP association and airtime control algorithm, and also the MAC parameter control algorithm. More specifically, the performance of our GP-based association scheme is compared with the Max-SNR scheme in terms of throughput and fairness. We implemented a simulator for EDCA (including access probability q and inter-frame space L described in Section VII-A) in Matlab to measure the achieved throughput numerically. Furthermore, we used cvx to solve the GP problems in the association algorithm.

We consider a network in which 4 APs are deployed in a $10 \times 10\ m^2$ area. More specifically, the APs are placed at the centers of four different $5 \times 5\ m^2$ grids to provide seamless coverage. To eliminate interference between the transmission of different APs, four non-overlapping 20 MHz channels are assigned to four APs. The STAs are distributed in the entire area according to the two-dimensional Poisson point process (PPP).

The wireless channel model includes path loss and small-scale fading. Generally, the channel gain can be expressed as $h = Ah'd^{-\alpha/2}$, where d is the distance between a STA and an AP, $\alpha \geq 2$ is the path loss exponent, A is a constant dependent on the frequency and transmitter/receiver antenna gain, and h' represents the small scale fading component. In the numerical results, we set $\alpha = 3$ and $A = 1$. Furthermore, h' is randomly generated according to the Rayleigh distribution assuming $\mathbb{E}\{|h'|^2\} = 1$. The received SNR at STA i is equal to $\frac{Pg_i^a}{\sigma^2}$ where P is the transmission power, $g_i^a = |h_i^a|^2$ is the channel power gain from STA i to AP a, and σ^2 is the power of noise. In all the numerical results, P/σ^2 is assumed fixed and set to 10dB.

To determine the transmission rate of each STA-AP pair, the 802.11a physical layer model is used. More specifically, to guarantee a maximum packet error rate, adaptive modulation

time-slot δ	9 μs
Propagation Delay γ	1 μs
T_{TXOP}	1 ms
SIFS	10 μs
ACK	40 μs

TABLE II: IEEE 802.11e MAC parameters used in the numerical results
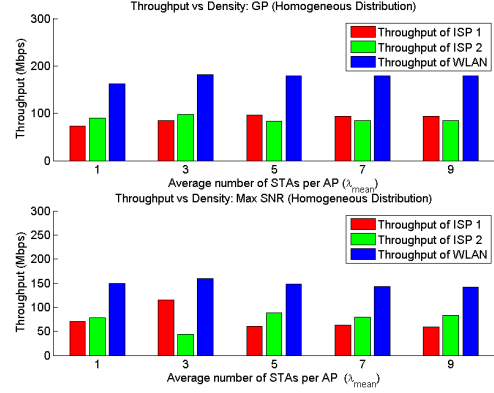

Fig. 7: Throughput vs. STA density, homogeneous distribution
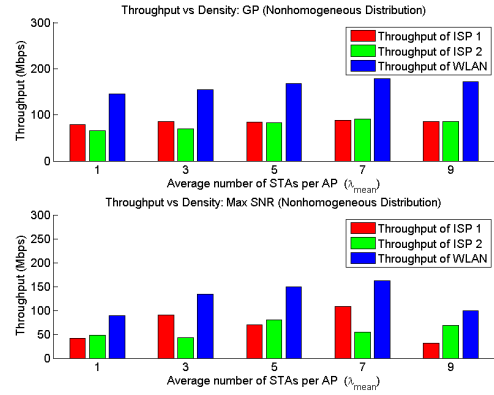

Fig. 8: Throughput vs. STA density, non-homogeneous distribution

and coding is used based on the received SNR. Table I shows the achievable transmission rates and adaptive modulation and coding schemes standardized in IEEE 802.11a and the SNR range for each scheme used in the simulations.

The MAC layer parameters used in our simulations are summarized in Table II. Moreover, the target airtime share for each ISP k (i.e., η_k) is set equal to the number of APs divided by the number of ISPs. In other words, we assume that the ISPs have the same minimum airtime reservation and share the total airtime in a fair manner. In the Markov chain, the average frozen time N is approximated by T_{TXOP}/δ.

A. Effects of STA Distribution

Here, we investigate the impact of STA distribution on the fairness and throughput achieved by the two association algorithms. More specifically, we set up two examples considering homogeneous and non-homogeneous STA distributions. In both examples, the STAs are randomly associated to the two ISPs with equal probabilities.
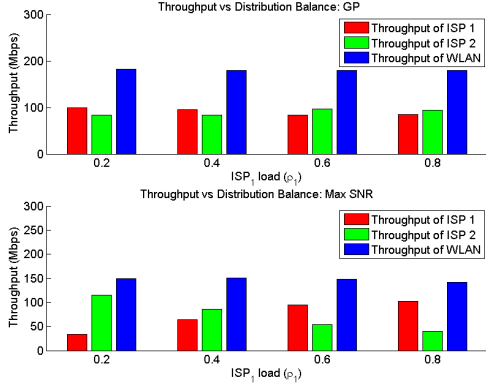
Fig. 9: Throughput vs. ISP 1 load

1) Homogeneous Distribution: In this example, the STAs are distributed in the square space according to a homogeneous two-dimensional PPP. Accordingly, the number of STAs in each grid (where an AP centered at) follows a Poisson distribution with mean λ_{mean}, which represents the average number of STAs per AP. Figure 7 shows that GP-based association scheme improves both fairness and total throughput compared with Max-SNR scheme.

2) Non-homogeneous Distribution: Let consider that STAs are distributed according to a non-homogeneous PPP. More specifically, the number of STAs in a grid (where AP a centered at) follows a Poisson distribution with mean λ_a. Here, λ_a is randomly generated between 0 and λ_{mean}. With such STA distribution, Figure 8 shows the achieved throughput of two ISPs versus λ_{mean}. As expected, the performance gap is significantly larger with non-homogeneous distribution compared to the homogeneous case. Max-SNR can hardly guarantee the fairness between different ISPs. While for the same distribution, GP-based STA-AP association can manage to keep the balance between two ISPs. Furthermore, it can offer improvement over Max-SNR in terms of the total throughput due to the load-balancing among APs.

B. Effects of STA Density and ISP Load

Let define ρ_1 (also referred to as ISP 1 load) as the ratio of number of STAs serving by ISP 1 to the total number of STAs in the network. Here, the performance of the two association approaches are compared under different STA density and ISP load.

Assuming a homogeneous STA distribution with $\lambda_{\mathrm{mean}} = 3$, Figure 9 demonstrates the achieved throughput of two ISPs versus different values of ρ_1 for both GP-based and Max-SNR association schemes. By Max-SNR association, it is shown that throughput of ISP 1 grows linearly with ρ_1, while the achieved throughput of ISP 2 is decreasing. But, GP-based association can fairly distribute the airtime between ISPs regardless of their ISP loads and thus keep a perfect balance between the achieved throughput of the two ISPs.

Figure 10 shows the total throughput achieved by the two association algorithms versus λ_{mean} for a homogeneous STA distribution. For a fixed ρ_1, the total throughput by both
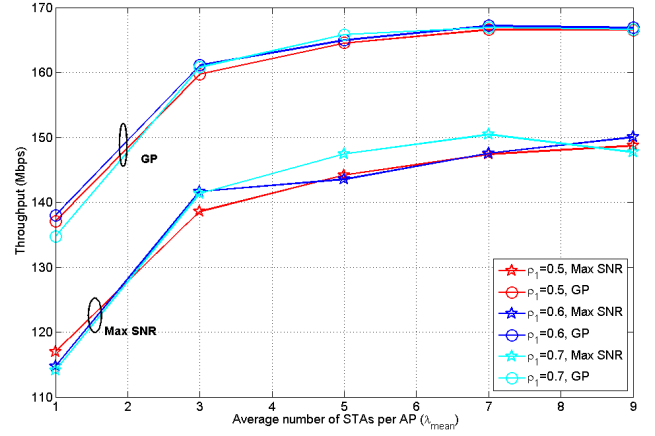

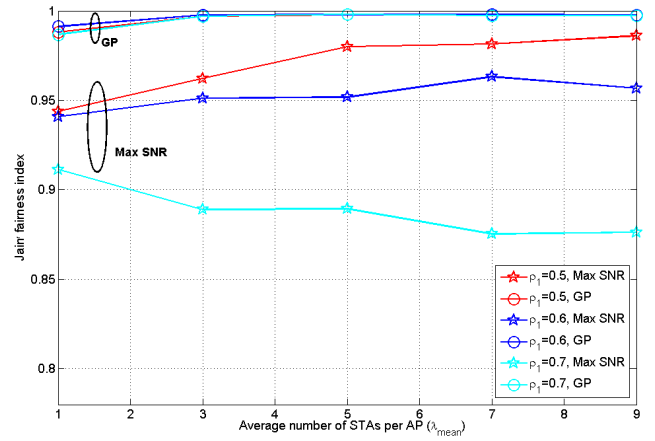Fig. 10: Total throughput vs. STA density for different ρ_1

Fig. 11: Fairness vs. STA density for different ρ_1

algorithms increases with the STA density (i.e., λ_{mean}). But, the throughput increase rate is decreasing with λ_{mean}. This is because the wireless channel is underutilized when the STA density is low. Thus, the increase in the STA density will improve the total throughput. But, when the STA density is large, increasing the STA density further will result in a higher collision probability, and hence, slow down the total throughput improvement. For any fixed ρ_1, it is shown that GP-based association significantly improves the total throughput compared with the Max-SNR association.

Figure 11 measures the fairness by employing the Jain's fairness index in (25).

$$F = \frac{(\sum_{k \in \mathcal{K}} T_k)^2}{|\mathcal{K}| \sum_{k \in \mathcal{K}} T_k^2} \tag{25}$$

where $T_k = \sum_{i \in S_k, a \in A} T_i^a$ is the achieved throughput for all the STAs of ISP k. From Figure 11, it is clear that the proposed GP-based association approach can always guarantee perfect fairness between the ISPs regardless of the STA density or ρ_1. The achieved fairness level by Max-SNR association is always worse than GP-based, especially when the STA load is highly unbalanced between ISPs (i.e., ρ_1 is not close to 0.5).

VI. CONCLUSION

This paper considers the STA-AP association and airtime control in virtualized 802.11 networks aiming to provide fairness guarantees among ISPs despite the number of STAs per ISP. First, a three-dimensional Markov chain is developed to model a generalized 802.11e EDCA protocol. This model establishes the relationship between the transmission probability of each STA and the detailed parameters in the MAC protocol. Based on this relationship, the feasibility region of the transmission probabilities are derived. Subsequently, an optimization problem is formulated which can maximize the network throughput, while guaranteeing the fairness between different ISPs. The implementation of the optimal transmission probabilities obtained by successive geometric programming is discussed by controlling the MAC parameters. Extensive numerical results confirm that the Markov chain can accurately describe the MAC protocol. Furthermore, it is verified that the proposed association algorithm can improve the throughput and provide fairness guarantees in virtualized 802.11 WLANs.

VII. APPENDIX

A. Markov Chain Model

Here, we detail the Markov chain model for EDCA introduced in Section III-A. Firstly, the transition rules among the states of the Markov chain are given according to the EDCA protocol. Then, the transition probability between any two adjacent states are summarized. Finally, by solving the stationary point of the Markov chain, the relationship between transmission probability and the detailed EDCA parameters are derived.

1) Transition rules: First, we describe the transitions among the states in the proposed Markov chain according to the EDCA protocol. As shown in Figure 4, after a successful transmission or a packet drop, the STA will flip a biased coin with successful probability q_i. If the coin appears to be tail (with probability $1 - q_i$), the STA will go to the state $\{-2, 0, L_i - 1\}$, and try again after L_i time-slots. Otherwise, if the coin is head (with probability q_i), the STA will move to the state $\{-1, 0, A_i\}$, where $A_i = \text{AIFS}[\text{STA } i] - 1$. If the channel stays idle in the following $A_i + 1$ time-slots, the STA will move on to the backoff/retransmission stage 0 and uniformly pick a random backoff time from $[0, W_{0,i}]$.

Otherwise, if the STA senses the channel busy in a state $\{-1, 0, d\}$ where $0 \le d \le A_i$, it will freeze and move to the state $\{-1, 0, N + A_i\}$, where N is the average frozen time (The frozen time is equal to either T_s or T_c and thus N can be approximated to T as discussed before). Afterwards, it will wait until channel becomes idle again, by counting down $v(t)$ from $N + A_i$ to A_i. When the STA gets to the state $\{-1, 0, A_i\}$, it has to keep repeating the explained process until it enters the backoff process. We assume that the probability that the channel is observed busy by STA i is a constant p_i in a general time-slot. Given the transmission

probabilities of all STAs (i.e., τ_i) in the BSS, we have

$$p_i = 1 - \prod_{i' \ne i} (1 - \tau_{i'}). \qquad (26)$$

For the backoff process, we use Figure 12 to explain the possible transitions at the backoff/retransmission stage j. When STA i reaches the state $\{j-1, 0, 0\}$, it will start a transmission. Then, if other STAs happen to access the channel at the same time, STA i will experience a collision. The probability that STA i encounters a collision is equal to p_i. In a case of collision, the STA will enter the next backoff/retransmission stage (i.e., j), pick a random backoff counter b between 0 and $W_{j,i}$, and go to the state $\{j, b, 0\}$. When the STA reaches the state $\{j, b, 0\}$, it will count down to $\{j, b-1, 0\}$ if the channel is idle. Otherwise, if the channel is sensed busy (i.e., other STAs are transmitting in the channel), STA i will freeze its backoff counter and move to the state $\{j, b, N + A_i\}$. In the frozen period, the STA will count down $v(t)$ by 1 each time-slot until it gets to the state $\{j, b, A_i\}$. Then, if the channel is still idle (with probability $(1 - p_i)$), the STA will continue counting down to $\{j, b, 1\}$ and then go to state $\{j, b - 1, 0\}$. Otherwise, if the channel becomes busy when STA i is in the state $\{j, b, d\}$ where $1 \le d \le A_i$, it has to freeze its backoff counter again and go back to the state $\{j, b, N + A_i\}$.

Finally, for states $\{j, 0, 0\}$, $0 \le j \le m_i + h_i$, the STA will initiate a transmission. If no other STA transmits at the same time (with probability $(1 - p_i)$), the STA will have a successful transmission of duration T_s. Otherwise, if any other STA simultaneously starts a transmission (with probability p_i), STA i will experience a collision of duration T_c. After the collision, if $0 \le j \le m_i + h_i - 1$, the STA will go to the next backoff/retransmission stage. But, if $j = m_i + h_i$ (which means that the maximum backoff/retransmission limit is reached), the STA has to drop this packet. Then, for a new packet transmission, it has to flip a biased coin and start over from the state $\{-2, 0, L_i - 1\}$ or $\{-1, 0, A_i\}$ according to the result of the coin toss. The mathematical representations of the transition probabilities are summarized in Section VII-A2.

2) Transition Probabilities of the Markov Chain: Given the transition rule between the states described in above, we have the following transition probabilities.

1) For states $\{-2, 0, d\}$, $0 \le d \le L_i - 1$, i.e., after getting a tail in a coin toss,

$$\begin{cases} P\{(-2, 0, d-1)|(-2, 0, d)\} = 1, \ 1 \le d \le L_i - 1 \\ P\{(-2, 0, L_i - 1)|(-2, 0, 0)\} = 1 - q_i \\ P\{(-1, 0, A_i)|(-2, 0, 0)\} = q_i \end{cases}$$

2) For states $\{-1, 0, d\}$, $0 \le d \le N + A_i$, i.e., before entering the backoff process,

$$\begin{cases} P\{(-1, 0, d-1)|(-1, 0, d)\} = 1, \ A_i + 1 \le d \le N + A_i \\ P\{(-1, 0, d-1)|(-1, 0, d)\} = 1 - p_i, \ 1 \le d \le A_i \\ P\{(-1, 0, N + A_i)|(-1, 0, d)\} = p_i, \ 0 \le d \le A_i \\ P\{(0, b, 0)|(-1, 0, 0)\} = \frac{1-p_i}{W_{0,i}+1}, \ 0 \le b \le W_{0,i} \end{cases}$$
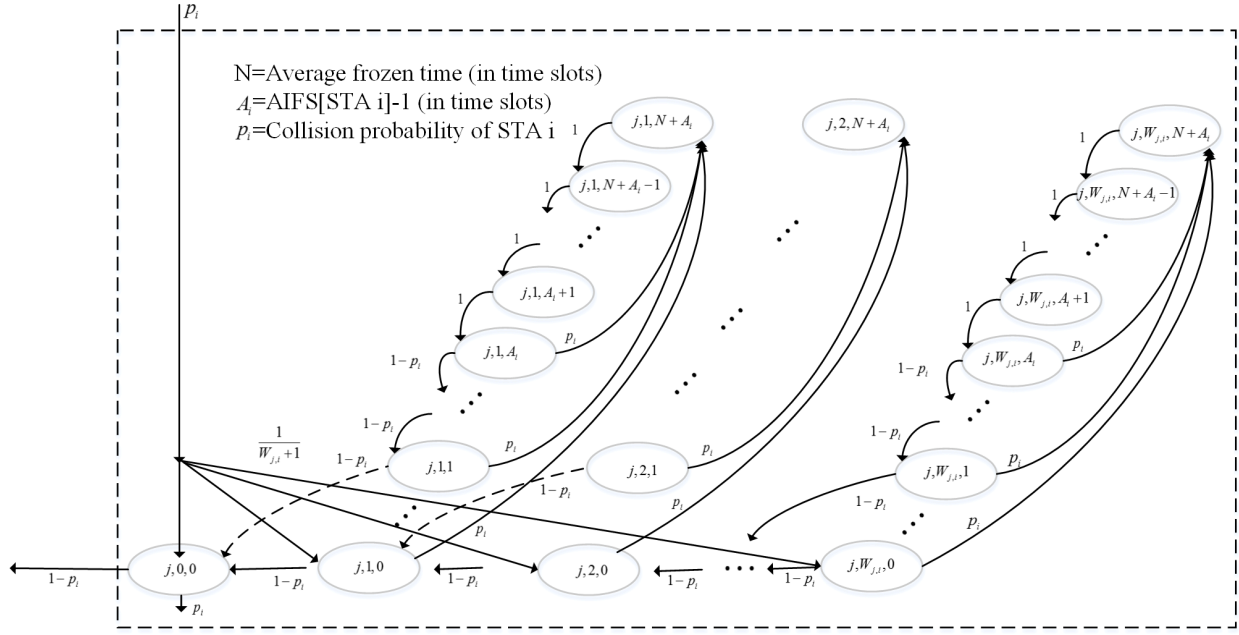
Fig. 12: Detailed transition for one backoff stage

3) For states $\{j, b, d\}$, $0 \le j \le m_i + h_i$, i.e., in the backoff process,

a) If $b = 0, d = 0$, i.e., before a successful transmission or a collision,

$$\begin{cases} P\{(-1, 0, A_i)|(j, 0, 0)\} = (1 - p_i)q_i, \\ \quad 0 \le j \le m_i + h_i - 1 \\ P\{(-2, 0, L_i - 1)|(j, 0, 0)\} = (1 - p_i)(1 - q_i), \\ \quad 0 \le j \le m_i + h_i - 1 \\ P\{(j + 1, b, 0)|(j, 0, 0)\} = p_i \frac{1}{W_{j+1, i} + 1}, \\ \quad 0 \le j \le m_i + h_i - 1, \ 0 \le b \le W_{j+1, i} \\ P\{(-2, 0, L_i - 1)|(m_i + h_i, 0, 0)\} = 1 - q_i \\ P\{(-1, 0, A_i)|(m_i + h_i, 0, 0)\} = q_i \end{cases}$$

b) If $d = 0, 1 \le b \le W_{j,i}$, i.e, in the count down process,

$$\begin{cases} P\{(j, b - 1, 0)|(j, b, 0)\} = 1 - p_i, \ 1 \le b \le W_{j,i} \\ P\{(j, b, N + A_i)|(j, b, 0)\} = p_i, \ 1 \le b \le W_{j,i} \end{cases}$$

4) If $1 \le d \le N + A_i$, $1 \le b \le W_{j,i}$, i.e., during a frozen period or the AIFS after it,

$$\begin{cases} P\{(j, b, d - 1)|(j, b, d)\} = 1, \ A_i + 1 \le d \le N + A_i \\ P\{(j, b, N + A_i)|(j, b, d)\} = p_i, \ 1 \le d \le A_i \\ P\{(j, b, d - 1)|(j, b, d)\} = 1 - p_i, \ 2 \le d \le A_i \\ P\{(j, b - 1, 0)|(j, b, 1)\} = 1 - p_i \end{cases}$$

3) Stationary Condition of the Markov Chain: Based on the transition rule above and the structure of the Markov chain, as in [31] and [30], we have,

$$b_{j,0,0} = p_i^j b_{0,0,0}, \ 0 \le j \le m_i + h_i$$

$$b_{j,b,0} = \frac{W_{j,i} + 1 - b}{W_{j,i} + 1} b_{j,0,0}, \ 0 \le j \le m_i + h_i, 1 \le b \le W_{j,i}$$

By exploiting the transition structure within a frozen process and the post AIFS (i.e., the states $\{j, b, d\}$, $0 \le d \le N + A_i$), we have,

$$b_{j,b,d} = \frac{p_i}{(1 - p_i)^d} b_{j,b,0}, 0 \le j \le m_i + h_i - 1, 1 \le d \le A_i - 1$$

$$b_{j,b,d} = \frac{p_i}{(1 - p_i)^{A_i}} b_{j,b,0}, 0 \le j \le m_i + h_i, A_i \le d \le N + A_i$$

For states $\{-2, 0, d\}$, by looking at the incoming and outgoing probability, we can have

$$b_{-2,0,d} = \frac{1 - q_i}{q_i} b_{0,0,0}, \ 0 \le d \le L_i - 1$$

For states $\{-1, 0, d\}$, by analyzing the structure of transition between the states $\{-1, 0, d\}, 0 \le d \le N + A_i$ and the state $\{0, 0, 0\}$, we can have

$$b_{-1,0,d} = \frac{1}{(1 - p_i)^{d+1}} b_{0,0,0}, \ 0 \le d \le A_i$$

$$b_{-1,0,d} = \frac{1 - (1 - p_i)^{A_i + 1}}{(1 - p_i)^{A_i + 1}} b_{0,0,0}, \ A_i + 1 \le d \le N + A_i$$

To this point, all the stationary probabilities in the Markov chain are represented in terms of $b_{0,0,0}$. Then, $b_{0,0,0}$ can be derived from the normalization condition in (27). Consequently, having all stationary probabilities $b_{j,b,d}$, the transmission probability of STA i can be calculated as in (13). To derive (13), the contention window $W_{j,i}$ is set equal to

$$W_{j,i} = \begin{cases} W_{\min,i} 2^j, & 0 \le j \le m_i \\ W_{\min,i} 2^{m_i}, & m_i + 1 \le j \le m_i + h_i \end{cases}$$

according to the exponential backoff rules in 802.11e. From the definition of transmission probability, τ_i can be calculated by summing all the stationary probabilities of the states in which the STA will initiate a transmission (i.e., $b_{j,0,0}, \ \forall j \in$

$$1 = \sum_{d=0}^{L_i-1} b_{-2,0,d} + \sum_{d=0}^{N+A_i} b_{-1,0,d} + \sum_{j=0}^{m_i+h_i} b_{0,0,0} + \sum_{j=0}^{m_i+h_i} \sum_{b=1}^{W_{j,i}} \sum_{d=0}^{N+A_i} b_{j,b,d}$$

$$= \left[L_i \frac{1-q_i}{q_i} + \frac{1+p_i N}{p_i} \frac{1-(1-p_i)^{A_i+1}}{(1-p_i)^{A_i+1}} + \frac{1-p_i^{m_i+h_i+1}}{1-p_i} + \frac{1+Np_i}{2(1-p_i)^{A_i}} \sum_{j=0}^{m_i+h_i} W_{j,i} p_i^j \right] b_{0,0,0} \qquad (27)$$

$\{0, 1, ..., m_i + h_i\}$). The expression in (13) establishes the relationship between τ_i and all the detailed parameters for STA i in the EDCA protocol, which will be useful to characterize the feasibility region of τ_i^a in the optimization and design an algorithm to control the transmission probabilities of the STAs.

B. Geometric Programming and Monomial Approximation

An optimization problem is called geometric programming if it follows the following form,

$$\min_{\boldsymbol{x}} f_0(\boldsymbol{x}), \quad \text{subject to}$$
$$f_i(\boldsymbol{x}) \leq 1, \; i = 1, \ldots, n_1$$
$$g_i(\boldsymbol{x}) = 1, \; i = 1, \ldots, n_2$$

where $f_0, \ldots, f_{n_1}$ are posynomials and $g_1, \ldots, g_{n_2}$ are monomials. In the context of geometric programming, a monomial function f of $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ is defined as,

$$f(\boldsymbol{x}) = c x_1^{a_1} x_2^{a_2} \ldots x_n^{a_n}$$

where $c > 0$ and $a_i \in \mathcal{R}$. Furthermore, a posynomial is defined as the summation of multiple monomials, i.e.,

$$g(\boldsymbol{x}) = \sum_{k=1}^{K} f_k(\boldsymbol{x})$$

The basic idea of monomial approximation is as follows: consider a posynomial function $g(\boldsymbol{x}) = \sum_k f_k(\boldsymbol{x})$ with $f_k(\boldsymbol{x})$ being the monomial terms. By the arithmetic-geometric mean inequality, we have

$$g(\boldsymbol{x}) \geq \hat{g}(\boldsymbol{x}) = \prod_k \left(\frac{f_k(\boldsymbol{x})}{\alpha_k(\boldsymbol{x}_0)} \right)^{\alpha_k(\boldsymbol{x}_0)} \qquad (29)$$

where the parameters $\alpha_k(\boldsymbol{x}_0)$ can be obtained by computing

$$\alpha_k(\boldsymbol{x}_0) = \frac{f_k(\boldsymbol{x}_0)}{g(\boldsymbol{x}_0)}, \; \forall k \qquad (30)$$

where $\boldsymbol{x}_0 > 0$ is a fixed point (e.g., the optimal solution from the last round of optimization). It is proved that $\hat{g}(\boldsymbol{x})$ is the best local monomial approximation of $g(\boldsymbol{x})$ near $\boldsymbol{x}_0$ [38].

REFERENCES

[1] A. Checco and D. Leith, "Fair virtualization of 802.11 networks," *IEEE/ACM Trans. Netw.*, vol. 23, no. 1, pp. 148–160, Feb 2015.

[2] M. Balazinska and P. Castro, "Characterizing mobility and network usage in a corporate wireless local-area network," in *Proc. of the 1st Intl. Conf. on Mobile Systems, Applications and Services.*, May 2003, pp. 303–316.

[3] D. Schwab and R. Bunt, "Characterising the use of a campus wireless network," in *Proc. IEEE Intl. Conf. on Computer Commun. (INFOCOM)*, Mar. 2004, pp. 862–870.

[4] H. Gong, K. Nahm, and J. W. Kim, "Distributed fair access point selection for multi-rate IEEE 802.11 WLANs," in *Proc. IEEE Consum. Commun. and Netw. Conf.*, Jan. 2008, pp. 528–532.

[5] O. Ercetin, "Association games in IEEE 802.11 wireless local area networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 5136–5143, Dec. 2008.

[6] W. Xu, C. Hua, and A. Huang, "A game theoretical approach for load balancing user association in 802.11 wireless networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2010, pp. 1–5.

[7] B. Pradeepa and J. Kuri, "An estimated delay based association policy for web browsing in a multirate WLAN," *IEEE Trans. Netw. Service Manag.*, vol. 9, no. 3, pp. 346–358, Sept. 2012.

[8] M. Heusse, F. Rousseau, G. Berger-Sabbatel, and A. Duda, "Performance anomaly of 802.11b," in *Proc. IEEE Intl. Conf. on Computer Commun. (INFOCOM)*, Mar. 2003, pp. 836–843.

[9] A. Checco and D. Leith, "Proportional fairness in 802.11 wireless LANs," *IEEE Commun. Lett.*, vol. 15, no. 8, pp. 807–809, Aug. 2011.

[10] X. Chen and D. Leith, "Proportional fair coding for 802.11 WLANs," *IEEE Commun. Lett.*, vol. 1, no. 5, pp. 468–471, Oct. 2012.

[11] S.-C. Liew and Y. Zhang, "Proportional fairness in multi-channel multi-rate wireless networks - Part I: The case of deterministic channels with application to AP association problem in large-scale WLAN," *IEEE Trans. Wireless Commun.*, vol. 7, no. 9, pp. 3446–3456, Sept. 2008.

[12] Y. Zhang and S.-C. Liew, "Proportional fairness in multi-channel multi-rate wireless networks - Part II: The case of time-varying channels with application to OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 7, no. 9, pp. 3457–3467, Sept. 2008.

[13] H. Zhou, P. Fan, and J. Li, "Global proportional fair scheduling for networks with multiple base stations," *IEEE Trans. Veh. Technol.*, vol. 60, no. 4, pp. 1867–1879, May 2011.

[14] T. Joshi, A. Mukherjee, Y. Yoo, and D. Agrawal, "Airtime fairness for IEEE 802.11 multirate networks," *IEEE Trans. Mobile Comput.*, vol. 7, no. 4, pp. 513–527, Apr. 2008.

[15] C.-T. Chou, K. Shin, and N. Sai Shankar, "Contention-based airtime usage control in multirate IEEE 802.11 wireless LANs," *IEEE/ACM Trans. Netw.*, vol. 14, no. 6, pp. 1179–1192, Dec. 2006.

[16] M. Laddomada, F. Mesiti, M. Mondin, and F. Daneshgaran, "On the throughput performance of multirate IEEE 802.11 networks with variable-loaded stations: analysis, modeling, and a novel proportional fairness criterion," *Wireless Communications, IEEE Transactions on*, vol. 9, no. 5, pp. 1594–1607, May 2010.

[17] P. Lin, W. i Chou, and T. Lin, "Achieving airtime fairness of delay-sensitive applications in multirate IEEE 802.11 wireless LANs," *IEEE Commun. Mag.*, vol. 49, no. 9, pp. 169–175, Sept. 2011.

[18] K. Nakauchi, Y. Shoji, and N. Nishinaga, "Airtime-based resource control in wireless LANs for wireless network virtualization," in *Fourth Intl. Conf. on Ubiquitous and Future Networks (ICUFN),*, July 2012, pp. 166–169.

[19] A. Banchs, P. Serrano, P. Patras, and M. Natkaniec, "Providing throughput and fairness guarantees in virtualized WLANs through control theory," *Mobile Networks and Applications.*, vol. 17, no. 4, pp. 435–446, June 2012.

[20] V. Ramaiyan, A. Kumar, and E. Altman, "Fixed point analysis of single cell IEEE 802.11e WLANs: Uniqueness and multistability," *IEEE/ACM Trans. Netw.*, vol. 16, no. 5, pp. 1080–1093, Oct 2008.

[21] I. Tinnirello and G. Bianchi, "Rethinking the IEEE 802.11e EDCA performance modeling methodology," *IEEE/ACM Trans. Netw.*, vol. 18, no. 2, pp. 540–553, Apr 2010.

[22] J. Robinson and T. Randhawa, "Saturation throughput analysis of IEEE 802.11e enhanced distributed coordination function," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 5, pp. 917–928, June 2004.

[23] A. Banchs and L. Vollero, "Throughput analysis and optimal configuration of 802.11 e EDCA," *Computer Networks*, vol. 50, no. 11, pp. 1749–1768, 2006.

[24] B. M. Parker, J. A. Schormans, and S. G. Gilmour, "Increasing throughput in IEEE 802.11 by optimal selection of backoff parameters," *IET Networks*, vol. 4, no. 1, pp. 21–29, 2014.

[25] P. Dong, J. Wang, H. Wang, and Y. Pan, "Boosting VoIP capacity via service differentiation in IEEE 802.11 e EDCA networks," *International Journal of Distributed Sensor Networks*, vol. 2015, 2015.

[26] J. Y. Lee, H. S. Lee, and J. S. Ma, "Model-based QoS parameter control for IEEE 802.11e EDCA," *IEEE Trans. Commun.*, vol. 57, no. 7, pp. 1914–1918, July 2009.

[27] Y. Gao, X. Sun, and L. Dai, "Ieee 802.11e EDCA networks: Modeling, differentiation and optimization," *IEEE Trans. Wireless Commun.*, vol. 13, no. 7, pp. 3863–3879, July 2014.

[28] B. Li, R. Battiti, and Y. Fang, "Achieving optimal performance by using the IEEE 802.11 MAC protocol with service differentiation enhancements," *IEEE Trans. Veh. Technol.*, vol. 56, no. 3, pp. 1374–1387, May 2007.

[29] Y. Ge, J. C. Hou, and S. Choi, "An analytic study of tuning systems parameters in ieee 802.11 e enhanced distributed channel access," *Computer Networks*, vol. 51, no. 8, pp. 1955–1980, 2007.

[30] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 3, pp. 535–547, Mar. 2000.

[31] Z.-N. Kong, D. H. K. Tsang, B. Bensaou, and D. Gao, "Performance analysis of IEEE 802.11e contention-based channel access," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 10, pp. 2095–2106, Dec. 2004.

[32] Y. Xiao, "IEEE 802.11e: QoS provisioning at the MAC layer," *IEEE Wireless Commun.*, vol. 11, no. 3, pp. 72–79, Jun. 2004.

[33] M. Derakhshani and T. Le-Ngoc, "Cognitive MAC designs: Background," in *Cognitive MAC Designs for OSA Networks*. Springer International Publishing, 2014, pp. 15–31.

[34] ——, "Adaptive access control of CSMA/CA in wireless LANs for throughput improvement," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2013, pp. 2951–2955.

[35] M. Chiang, *Geometric programming for communication systems*. Now Publishers Hanover, 2005.

[36] G. Xu, "Global optimization of signomial geometric programming problems," *European Journal of Operational Research*, vol. 233, no. 3, pp. 500–510, 2014.

[37] M. Hershenson, S. Boyd, and T. Lee, "Optimal design of a cmos op-amp via geometric programming," *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*, vol. 20, no. 1, pp. 1–21, Jan 2001.

[38] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.